BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

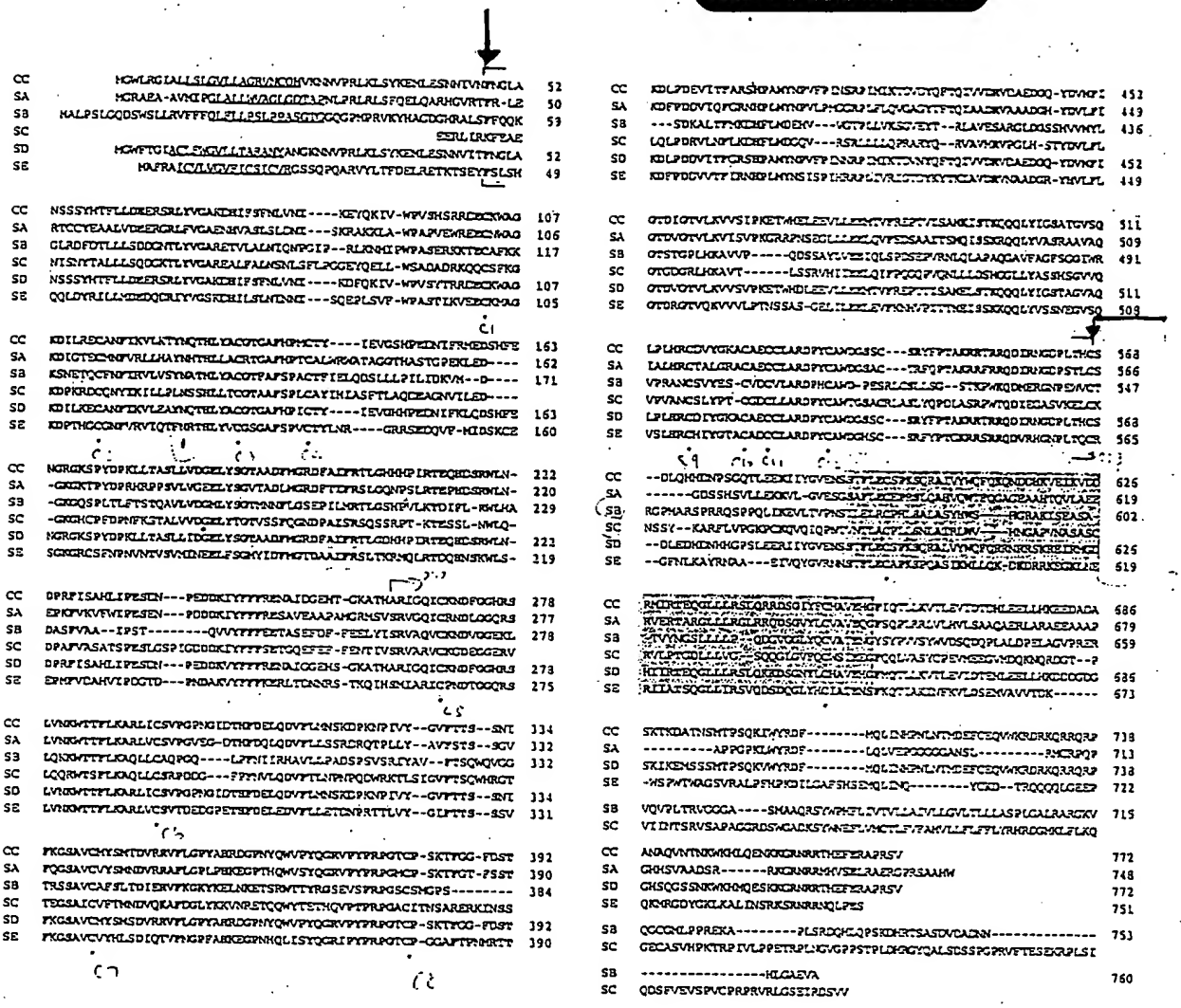


Figure 1. Alignment of Murine Semaphorin Sequences

The sequences of Sem A (SA), Sem B (SB), Sem C (SC), Sem D (SD), and Sem E (SE) were aligned with that of chick collapsin (CC) using the programs CLUSTAL and PILEUP (HUSAR 3.0 software). Putative signal sequences determined with the program SIGNALSEQ (HUSAR 3.0 software) are underlined, conserved cysteines are indicated by asterisks, and a conserved potential N-glycosylation site is marked by a number symbol. Conserved residues present in more than half of the sequences (bold letters), semaphorin domains (brackets), and immunoglobulin homologies (gray box) are indicated. Several partial cDNAs were isolated for Sem A, Sem B, Sem C, Sem D, and Sem E, and one clone containing the complete coding sequence for Sem B. Both strands of the longest cDNAs were sequenced for each semaphorin. The full coding sequences for Sem A, Sem B, Sem D, and Sem E were determined from cDNA clones and PCR fragments amplified by rapid amplification of cDNA ends (RACE). No complete coding sequence was obtained for Sem C.

Two groups of murine semaphorin proteins can be distinguished (Figure 2). Three cDNAs (group III) *semA*, *semD*, and *semE*; also see Discussion) encode proteins that are highly homologous to the chicken collapsin and H-Sema III and include Sem D, which is likely to represent the murine collapsin homolog. The semaphorin domains of Sem A, Sem D, and Sem E are 53%–62% identical (70%–79% homologous) to each other. These proteins also share considerable homology throughout their entire coding sequence (45%–55% amino acid identity and 63%–73% homology). The second group (group IV; Sem B and Sem C) displays only about 50% homology (around 30% identity throughout the protein sequences, 33%–39% identity in the semaphorin domains) with other semaphorins, but shares most of the motifs characteristic of and a potential N-glycosylation site. Sem B and Sem C (group IV) are most closely related to each other (61% homology and 43% identity within the semaphorin domain), but to a lower extent than is found between the different members of the Sem D/collapsin group. Calculation of the phylogenetic relationships (Figure 2B) indicates that the vertebrate semaphorins form a separate group that, within the semaphorin domains, shows a maximal homology of 58% with the invertebrate semaphorin proteins.

Semaphorins Are Expressed Early in Embryonic Development

Northern blot hybridization of poly(A)$^+$ RNA prepared from

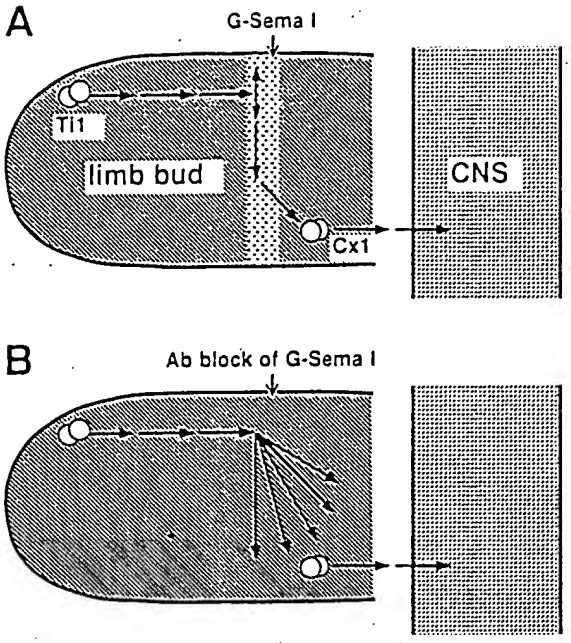


Figure 1. Expression and Function of Sema I in the Grasshopper Limb Bud

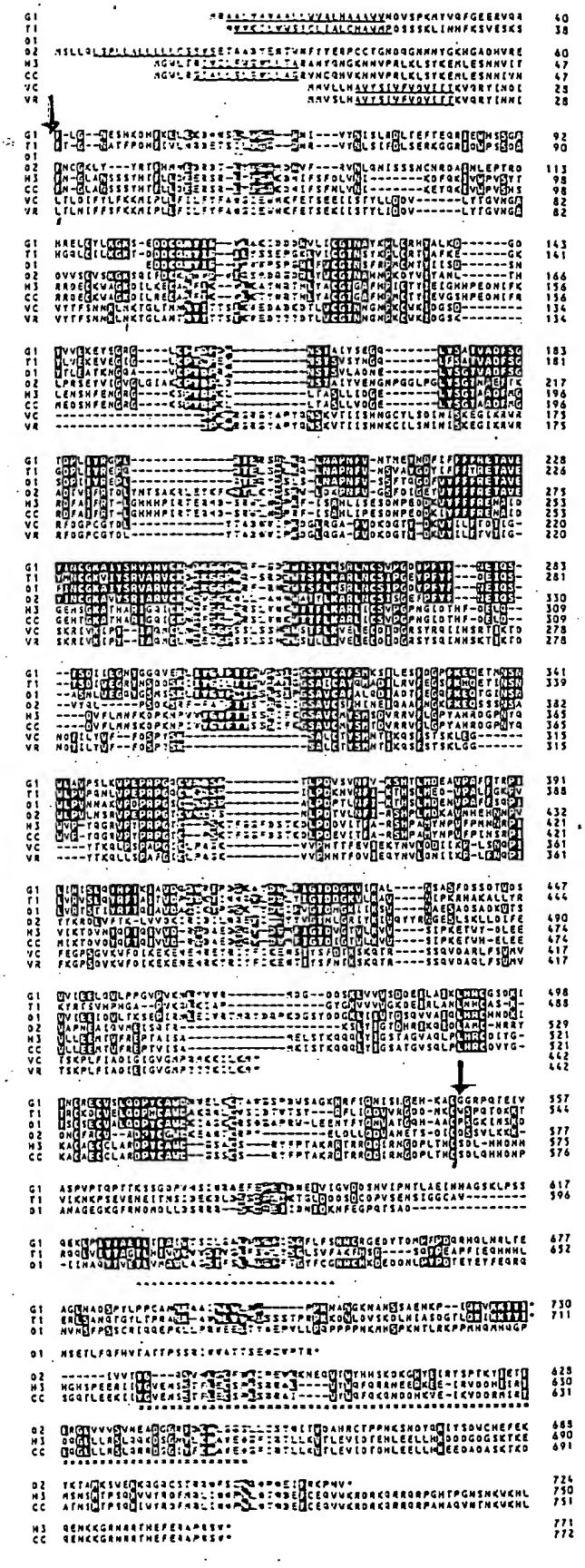
Schematic diagrams of 35% grasshopper limb buds showing (A) the trajectory of the pair of Ti1 growth cones in relationship to the epithelial expression of Sema I as they pioneer an axon pathway towards the CNS, and (B) the aberrant trajectory of these growth cones following application of antibodies that bind Sema I (Kolodkin et al., 1992).

(A) The long dark arrows represent the sequential steps in the pathfinding by the Ti1 growth cones (O'Connor et al., 1990). They first extend proximally towards the CNS, then stop and reorient once they encounter the Sema I boundary (initially extending in both the dorsal and ventral directions), then extend ventrally as a tightly fasciculated pair of axons along the distal portion of Sema I–expressing cells, reorient as they turn proximally after filopodial contact with the Cx1 neurons, and finally extend into the CNS.

(B) In cultured limb buds in which antibodies against Sema I were added during the period of axon outgrowth, the Ti1 growth cones often extend across the Sema I–expressing stripe, ultimately turning ventrally towards the Cx1 neurons, but doing so in a more variable fashion. The growth cones branched into multiple axons that were highly defasciculated in this location. See Discussion.

sophila melanogaster) diverged some 300 million years ago. Moreover, since G-Sema I was a pioneer protein with no similarity to other proteins and no internal repeats, it was difficult to predict which residues might be conserved across species.

A PCR-based evolutionary approach addressed this problem, incorporating the use of degenerate primers designed from several G-Sema I sequences that we guessed might be of structural or functional significance. One pair of primers (see Experimental Procedures) was used in low stringency PCR protocols with genomic DNA templates to isolate related sequences from a beetle, Tribolium confusum, and from a moth, Manduca sexta. The PCR product from Tribolium was used to isolate cDNA clones. The deduced amino acid sequence for the Tribolium open



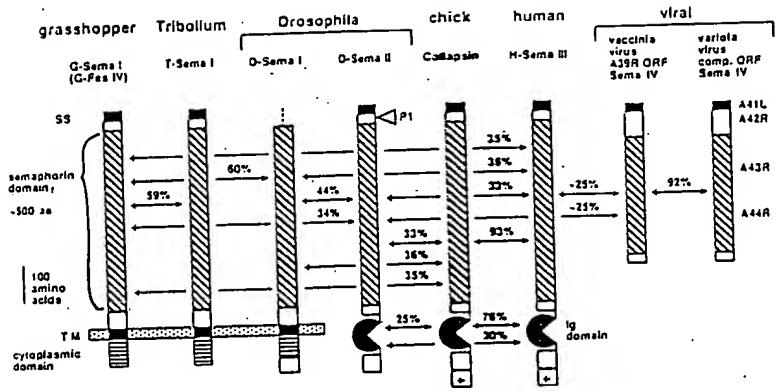


Figure 3. Schematic Diagram of Semaphorin Proteins

Schematic representations of G-Sema I (Kolodkin et al., 1992), the other three insect, one human, and two viral Sema proteins described here and chick collapsin (Luo et al., 1993). These proteins begin with a signal sequence (SS) and are thus likely to encode transmembrane or secreted proteins. All six eukaryotic sema share a 500 amino acid Sema domain (hatched boxes). The two viral sequences encode truncated and more divergent Sema domains. The percentages along the arrows denote the percent amino acid identity between the aligned domains in the pairs of sequences. Three of the insect proteins appear to be homologs (G-Sema I, T-Sema I, and D-Sema I); all three have a transmembrane domain (TM) followed by a modestly conserved cytoplasmic domain (horizontal stripes). D-Sema I has an additional 32 amino acids at the C-terminus of its cytoplasmic domain (open box) not found in the other two (the N-terminal sequence of D-Sema I is incomplete). D-Sema II, chick collapsin, and H-Sema III have neither transmembrane domains nor any other potential membrane linkage, but rather have a single Ig domain (closed semicircle) followed by another short stretch of amino acids (open box). Chick collapsin and H-Sema III have an additional 45 amino acids at their C-terminus that include 13–19 (respectively) positively charged amino acids; this region is shown as a basic (plus sign) domain. The vaccinia virus Sema sequence is found in ORF A39R. The variola virus Sema sequence was compiled from the nucleotide sequences in and around ORFs A41L, A42R, A43R, and A44R. D-Sema II has a triangle labeled P1 that denotes the location of a P element inserted in the codon for amino acid 32 that leads to the loss-of-function mutation *semaII*^{P1}.

information was then used to refine our primer design for PCR reactions using Drosophila embryonic cDNA.

Three pairs of primers (see Experimental Procedures) were used successfully to amplify Drosophila sequences related to G-Sema I. Two pairs of primers (that shared a common 3′ primer) gave products encoding G-Sema I–related sequences that are similar but not identical. Subsequent analysis confirmed the identification of two different Drosophila *sema* genes. The two different PCR products were used to isolate two different classes of cDNA clones. The two Drosophila ORFs are shown in Figure 2, and the protein structures are shown schematically in Figure 3.

The Tribolium protein and one of the Drosophila proteins appear to be homologs of the G-Sema I protein because they share 60% amino acid identity with G-Sema I along its entire length (see Figure 2; Figure 3). Thus we call these proteins Tribolium Sema (T-Sema) I and D-Sema I. The other Drosophila protein has a lower degree of amino acid identity and a distinct C-terminal structure, defining a second protein called D-Sema II.

The two new complete insect sequences (T-Sema I and D-Sema II) encode proteins with signal sequences (von Heijne, 1986), as does G-Sema I. T-Sema I and D-Sema I have transmembrane domains (Chou and Fasman, 1974), as does G-Sema I. However, D-Sema II has no transmembrane domain, its C-terminus shows no indication of a potential phospholipid linkage, and thus it is likely to be secreted. In addition, at its C-terminus, D-Sema II has a single C2-type immunoglobulin (Ig) domain (Williams and Barclay, 1988) (see Figure 2; Figure 3).

All four insect Semaphorins share a highly conserved extracellular domain of 500 amino acids (using basic local alignment search tool P. Altschul et al., 1990; the PIR data base) that is characterized by 16 conserved cysteines, one conserved potential N-linked glycosylation site, and numerous blocks of conserved amino acids throughout the 500 amino acid domain (see Figure 2). In their Semaphorin domains (comprising two-thirds of each protein), the three putative homologs (G-Sema I, T-Sema I, and D-Sema I) are most similar and all are more divergent from D-Sema II in terms of both percent identity and the absence of

Figure 2. Amino Acid Alignments of Semaphorin Proteins

Alignments of G-Sema I (G1) (Kolodkin et al., 1992), the other three insect (T1, D1, D2), one human (H3), and two viral (VC, VR) Sema amino acid sequences presented in this study, and chick collapsin (CC) (Luo et al., 1993), are shown. All of these proteins share a highly conserved 500 amino acid domain (delimited by the number symbols), that defines the Sema domain. Tribolium Sema I (T1) and D-sema I (D1) appear to encode homologs of G-sema I in that all three are highly conserved transmembrane proteins (the sequence for D-Sema I is incomplete in its N-terminal end). In contrast, D-Sema II (D2), H-Sema III (H3), and chick collapsin encode proteins that lack a transmembrane domain, have a single immunoglobulin domain, and have no apparent membrane linkage. H-Sema III and chick collapsin, but not D-Sema II, end with a highly basic domain. The two viral sequences (vaccinia virus A39R ORF [VC] and variola virus composite ORF [VR]) encode proteins with truncated and more divergent Sema domains (question mark denotes residues in the variola sequence that are ambiguous at points of likely frame shifts; see Results). In the Sema domain, we define conserved residues by their presence in at least three protein sequences, only one of which can be viral (owing to their 92% amino acid identity) and only one of which can be collapsin or H-Sema III (owing to their 93% amino acid identity). In the transmembrane and cytoplasmic domains, we define conserved residues by their presence in at least two of the three protein sequences. Black outline, conserved